Inference of Temporally Varying Bayesian Networks

Thomas Thorne & Michael P.H. Stumpf
Centre of Integrative Systems Biology and Bioinformatics,
Division of Molecular Biosciences, Imperial College London,
London SW7 2AZ, UK

February 24, 2013

Abstract

When analysing gene expression time series data an often overlooked but crucial aspect of the model is that the regulatory network structure may change over time. Whilst some approaches have addressed this problem previously in the literature, many are not well suited to the sequential nature of the data. Here we present a method that allows us to infer regulatory network structures that may vary between time points, utilising a set of hidden states that describe the network structure at a given time point. To model the distribution of the hidden states we have applied the Hierarchical Dirichlet Process Hidden Markov Model, a nonparametric extension of the traditional Hidden Markov Model, that does not require us to fix the number of hidden states in advance. We apply our method to existing microarray expression data as well as demonstrating is efficacy on simulated test data.

1 Introduction

The analysis of gene expression data in the field of systems biology is a challenging problem for a number of reasons, not least because of the high dimensionality of the data and relative dearth of data points. A number of approaches have been taken to inferring regulatory interactions from such data, often employing graphical models or sparse regression techniques [13, 18, 22].

These problems are further compounded by the nature of the biological systems under consideration, due to the influence of unobserved actors that may alter the behaviour of the system. Often experiments are performed over long time periods during which it is natural to expect the regulatory interactions at work to change. It is also worth noting that the time scales of regulatory responses to stimuli often differ from those of signalling and metabolic responses, and so it may be that responses to stimuli, around which experiments are often designed, take place in several phases each having different time scales.

Previous studies have attempted to address this problem by introducing changepoints in the time series, allowing the inferred network structure to differ between the resulting segments of the time series. For example in Lèbre *et al.* [14] a changepoint model is applied in which Dynamic Bayesian Networks are inferred for each segment of the time series. However such approaches may place strong prior assumptions on the number of changepoints that may be observed, and do not adjust for the complexity of the observed data automatically. Instead in Grzegorczyk *et al.* [10] an allocation sampler is used in combination with Bayesian Networks to assign each observation to a group, but unlike changepoint models this method treats the observations as being exchangeable, ignoring the fact that the data are sequential. The similar methodology in Ickstadt [11] uses a more flexible nonparametric prior on group assignments, applied to the modelling of molecular interactions using Bayesian Networks, but suffers the same drawbacks in not recognising the sequential nature of the data.

1

Here we present a methodology that allows us to infer network structures that may change between observations in a nonparametric framework whilst modelling the sequential nature of the data. To that end we have employed the infinite hidden Markov model (iHMM) of Beal *et al.* [1], also known as the hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) [25], in particular the "Sticky" extension of Fox *et al.* [5], in conjunction with a Bayesian network model of the gene regulatory network structure. The HDP-HMM allows the number of different states of the network structure to adapt as necessary to explain the observed data, including a potentially infinite number of states, of course restricted in practice by the finite number of experimental observations. In the previous work of Rodriguez *et al.* [20] it was demonstrated that the HDP-HMM outperforms a Dirichlet Process mixture for Gaussian graphical models on heterogeneous time series.

We apply our methodology to both simulated data and gene expression data for *Arabadopsis Thaliana* and *Drosophila Melanogaster*, demonstrating its effectiveness in detecting changes in network structure from time series data.

2 Methods

Given gene expression time series data over m genes at n time points, we denote the observations as the $n \times m$ matrix $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})^T$, where $\boldsymbol{x_j} = (x_{j1}, \ldots, x_{jm})^T$, the column vector of expression levels for each of the m genes at time point j. We formulate our model as a Hierarchical Dirichlet Process Hidden Markov Model, a stochastic process whereby a set of hidden states $s_1, \ldots, s_n$ govern the parameters of some emission distribution F over a sequence of time points $1 \ldots n$.

Each observation $\boldsymbol{x_j}$ is then generated from a corresponding emission distribution $F(\theta_k)$, where $s_j = k$. For our emission distributions, F, we use a Bayesian Network model over the m variables to represent the regulatory network structures corresponding to each hidden state.

2.1 Hierarchical Dirichlet Process Hidden Markov Models

To model a hidden state sequence that evolves over time we apply the methodology first introduced in Beal *et al.* [1] whereby a finite state Hidden Markov Model, consisting of a set of hidden states $s_1, \ldots, s_n$ over some alphabet $1 \ldots K$, is extended so that $K \to \infty$. In a classical Hidden Markov Model, the number of states K is typically specified in advance, and states follow a Markovian distribution whereby transitions are made between states with probability $\pi_{kl} = p(s_j = l | s_{j-1} = k)$, so that the next state in the sequence is drawn conditional only on the previous state.

The Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) [1, 25, 26] instead applies a Dirichlet Process prior to the transition probabilities $\boldsymbol{\pi}_k$. out of each of the states, and uses a hierarchical structure to couple the distributions between the individual states to ensure a shared set of potential states into which transitions can be made across all of the $\boldsymbol{\pi}$. This allows for an unlimited number of potential states, of course limited in practice by the number of observed data points.

More formally, each hidden state k possesses a Dirichlet Process G_k from which the next state is drawn, and a common (discrete) base measure G_0 is shared between the G_k, so that $G_k \sim \mathrm{DP}(\alpha, G_0)$. As a result transitions are made into a discrete set of states shared across all of the G_k, and drawn from G_0. The base measure G_0 is in turn drawn from a Dirichlet Process, $G_0 \sim \mathrm{DP}(\gamma, H)$, H being our prior over parameters for the emission distributions F_k.

Then using the stick breaking construction of Sethuraman [23] for G_0 and drawing θ_l independently from H, we have that $G_0 = \sum_l^\infty \beta_l \delta_{\theta_l}$, with $\boldsymbol{\beta} \sim GEM(\gamma)$, and so $G_k = \sum_l^\infty \pi_{kl} \delta_{\theta_l}$ with $\boldsymbol{\pi}_k \sim \mathrm{DP}(\alpha, \boldsymbol{\beta})$. The resulting model is outlined in figure 1.

However in a biological system it is probably realistic to assume that only a subset of the wide variety of potential behaviours of the hidden state sequence is relevant, as behaviour such as rapid
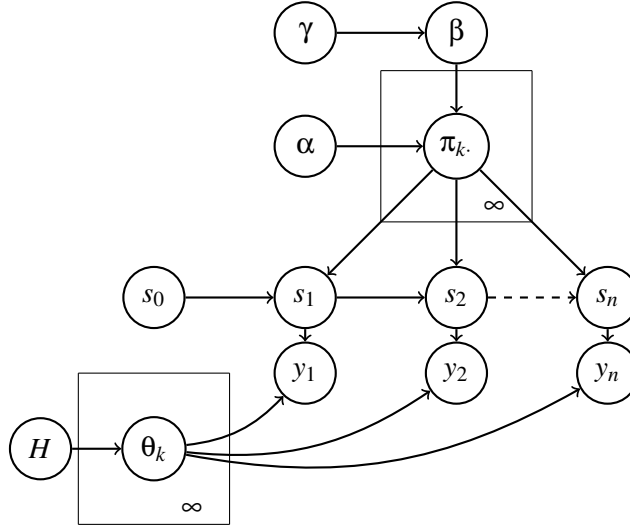
Figure 1: The Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM) represented as a graphical model.

cycling between states at adjacent time points would *a priori* seem to be unlikely to be observed in most gene expression data sets.

Thus we choose to apply the Sticky HDP-HMM [4, 5], that introduces an extra parameter κ that biases the prior probability of transitions between states towards remaining in the current state rather than transitioning to a differing one. Adding such a prior assumption simply states that we expect the state of the system to remain the same between successive time points; this is both parsimonious and would seem to be justified in the case of gene expression microarray data sets, where we might only expect to observe a small number of transitions to differing states across the time series.

This modification to the HDP-HMM gives us a model generating observed data points x_j as [5]

$$\boldsymbol{\beta}|\gamma \sim \text{GEM}(\gamma), \tag{1}$$

$$\boldsymbol{\pi_{k\cdot}}|\alpha, \boldsymbol{\beta}, \kappa \sim \text{DP}\left(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\delta_k}{\alpha + \kappa}\right), \tag{2}$$

$$s_j|s_{j-1}, \boldsymbol{\pi} \sim \boldsymbol{\pi}_{s_{j-1}\cdot}, \tag{3}$$

$$\theta \sim H, \tag{4}$$

$$x_j|s_j \sim F(\theta_{s_j}). \tag{5}$$

2.2 Gibbs sampling for the Sticky HDP-HMM

To sample from the hidden state sequence we have used a Gibbs sampling procedure based on the conditional probabilities for the hidden state s_i given the remaining hidden states s_{-i} [4], updating

each hidden state individually in a sweep over the n states,

$$
\begin{aligned}
p(s_j = k | \boldsymbol{s}_{-j}, \alpha, \beta, \kappa) \propto & (N^{-j}_{s_{j-1}k} + \alpha\beta_k + \kappa\delta_{s_{j-1}}(k)) \\
& \left(\frac{N^{-j}_{ks_{j+1}} + \alpha\beta_{s_{j+1}} + \kappa\delta_{s_{j+1}}(k) + \delta_{s_{j-1}}(k)\delta_{s_{j+1}}(k)}{\alpha + N^{-j}_{k\cdot} + \kappa + \delta_{s_{j-1}}(k)} \right) \\
& p(\boldsymbol{X}_{j\cdot} | \boldsymbol{X}_{i\cdot} : s_i = k, i \neq j, F_k),
\end{aligned} \tag{6}
$$

where $\boldsymbol{s}_{-j}$ denotes the state sequence $s_1, \ldots, s_n$ excluding s_j, N^{-j}_{kl} indicates the number of observed transitions from state k to state l within the hidden state sequence $\boldsymbol{s}_{-j}$, and $N^{-j}_{k\cdot}$ the total number of transitions from state k within $\boldsymbol{s}_{-j}$.

2.3 Bayesian Network emission distributions

To model the regulatory network structure corresponding to the hidden states of the HDP-HMM, we have applied a Bayesian Network methodology to capture the relationships between the genes represented in our data. Thus each hidden state has a unique Bayesian Network describing the interactions occurring between the genes at the time points corresponding to a particular state.

Bayesian Networks are probabilistic models, whereby a directed graph defines the conditional independence relationships between a set of random variables [12]. For the model to remain consistent, the graph structure $\mathcal{G}$, with nodes $u \in N_{\mathcal{G}}$ representing random variables and directed edges $(v, u) \in E_{\mathcal{G}}$ representing conditional probability relationships between them, must be acyclic.

For a given Bayesian network structure $\mathcal{G}$ and model parameters $\boldsymbol{\theta}$, the joint distribution $p(\boldsymbol{X}|\mathcal{G}, \boldsymbol{\theta})$ factorises as a product of local distributions for each node,

$$
p(\boldsymbol{X}|\mathcal{G}, \boldsymbol{\theta}) = \prod_{u \in N} p(\boldsymbol{x}_{\cdot u} | \mathrm{pa}_{\mathcal{G}}(u), \theta_u), \tag{7}
$$

where for each observation the value x_{iu} of a node u is dependent on the values of its set of parent nodes $\mathrm{pa}_{\mathcal{G}}(u) = \{v \in N | (v, u) \in \mathcal{G}\}$ and some parameters θ_u. Here we have used a BGe model [7] that allows the variables to take continuous values and defines the local distributions for each observation $i \in 1, \ldots, m$ of a gene u as

$$
x_{iu} \sim \mathcal{N} \left(\mu_u + \sum_{v \in \mathrm{pa}_{\mathcal{G}}(u)} b_{uv}(x_{iv} - \mu_v), \sigma_u^2 \right), \tag{8}
$$

with parameters $\theta_u = \mu_u, \boldsymbol{b_u}, \sigma_u^2$. Then using a conjugate Normal-Wishart prior and making certain assumptions we can calculate the local marginal likelihoods $p(\boldsymbol{x}_u | \mathrm{pa}_{\mathcal{G}}(u))$ as described in Geiger and Heckerman [7] and from these derive the joint probability $p(\boldsymbol{X}|\mathcal{G})$.

Unfortunately, due to the restriction of the network structure to that of a directed acyclic graph (DAG) it is difficult to explore the space of possible network structures. Several MCMC schemes have been proposed, including those of Grzegorczyk and Husmeier [9] and Madigan *et al.* [16], but performing random walks over DAG network structures faces the problem that proposing moves that maintain the DAG structure can be complex and time consuming, and mixing of the Markov chain can be slow.

However, as noted in Friedman and Koller [6], a DAG structure $\mathcal{G}$ corresponds to a partial ordering on the nodes and so induces a (non-unique) total ordering, and so the method of Friedman and Koller [6] instead proposes to perform a random walk over total orderings of the nodes. This allows the

Markov chain to move quickly around the space of possible graph structures, improving the mixing properties of the chain.

Whilst this introduces a bias in the prior distribution over graph structures [9] it greatly simplifies the computational complexity of the MCMC procedure and such a bias may be justified by arguments of parsimony, since graphs consistent with more orderings are more likely to be sampled. Furthermore the uniform prior on DAG structures is not uniform over Markov equivalent graphs, and so also introduces a different kind of bias in the results. Finally, a trivial modification of the algorithm of Friedman and Koller [6] allows for a correction of the bias [3].

Thus in our methodology we apply the MCMC sampler of Friedman and Koller [6] to infer Bayesian Network structures for each hidden state of the HDP-HMM by sampling over total orderings of the nodes $\sqsubset$ given the data points corresponding to the state in question. It is easy to calculate the likelihood of an ordering $\sqsubset$ using [6]

$$p(\boldsymbol{X}|\sqsubset) = \prod_{u \in N_{\mathcal{G}}} \sum_{k \in \mathrm{pa}_{\mathcal{G}}^{\sqsubset}(u)} p(\boldsymbol{x}_{\cdot u}, k), \tag{9}$$

where $\mathrm{pa}_{\mathcal{G}}^{\sqsubset}$ denotes the set of possible parent sets over the nodes of $\mathcal{G}$ consistent with the ordering $\sqsubset$. Then we can use a Metropolis Hastings sampler to sample from the posterior of orderings $p(\sqsubset|\boldsymbol{X}) = p(\boldsymbol{X}|\sqsubset)p(\sqsubset)$ [3], by beginning with an initial ordering and proposing and accepting new orderings $\sqsubset'$, distributed as $q(\sqsubset'|\sqsubset)$ with probability according to the Metropolis Hastings acceptance probability

$$p_{\mathrm{acc}} = \min\left(1, \frac{p(\boldsymbol{X}|\sqsubset')p(\sqsubset')q(\sqsubset|\sqsubset')}{p(\boldsymbol{X}|\sqsubset)p(\sqsubset)q(\sqsubset'|\sqsubset)}\right), \tag{10}$$

over a number of iterations. We choose to propose changes by swapping nodes in the ordering rather than more complex schemes such as "deck cutting", as these were found to have little impact on performance in previous studies [3, 6]. Proposals $\sqsubset' \sim q(\cdot|\sqsubset)$ are thus drawn by selecting two nodes within the ordering uniformly at random and exchanging their positions to produce a new ordering. In the absence of a compelling alternative, we take the prior over orderings $p(\sqsubset)$ as the uniform distribution.

Then for our emission distribution for a given state k we apply a Bayesian Network ordering $\sqsubset_k$ generating observed data points $\boldsymbol{X}^k$ distributed as $p(\boldsymbol{X^k}|\sqsubset_k)$ where by $\boldsymbol{X}^k$ we denote the subset of X_{ij} including only rows i corresponding to states $s_i = k$.

The full method is outlined in algorithm 1, and combines Gibbs updates of the hidden state sequence with Metropolis Hastings updates of the node orderings of the Bayesian Networks for each state at every iteration. To sample hidden state sequences and orderings from the posterior distribution the algorithm is first run for a number of burn-in iterations, after which samples are collected. Since a single iteration of our algorithm combines a full Gibbs update sweep along with an update of the Bayesian Network orderings over a number of Metropolis Hastings steps, in practice a comparatively small number of iterations of the algorithm are required to reach the posterior. To reduce the computational complexity of the Bayesian Network inference, we restrict the number of potential parents of a gene to be 2 or less. Even in such a case, we still face a large number of possible parent sets, of size $\sum_{i=2}^{m-1} \binom{i}{2} + \sum_{i=1}^{m-1} \binom{i}{1}$, and so in the analyses presented below we restrict our data set to a subset of genes of special interest, as is commonly the case in gene expression data analysis.

Finally, once we have inferred the hidden state sequence and generated a posterior sample of orderings corresponding to each state, we can then easily sample DAG structures from the posterior by first sampling an order from the posterior of a given state, and then sampling from the graphs consistent with this ordering, weighting the choice of parents by the local scores, and optionally attempting to account for the bias in the prior as described in Ellis and Wong [3].

Initialise state sequence $s_1, \ldots, s_n$;
Set $K \leftarrow$ number of states in s;
Initialise orderings $\sqsubset_1, \ldots, \sqsubset_K$;
for $iter \leftarrow 1$ **to** N_{iter} **do**
    for $j \leftarrow 1$ **to** n **do**
        for $k \leftarrow 1$ **to** K **do**
           $l_k \leftarrow p(\sqsubset_k | \boldsymbol{X}^k) p(s_j = k | s_{-j}, \alpha, \beta, \kappa)$;
        end
        Sample new ordering $\sqsubset_{K+1}$ from $p(\sqsubset)$ $l_{K+1} \leftarrow p(\sqsubset_{K+1} | \boldsymbol{x}_j) p(s_j = K + 1 | s_{-j}, \alpha, \beta, \kappa)$;
        $s_j \leftarrow$ sample from $1, \ldots, K + 1$ with weights $\boldsymbol{l}$;
        Update K;
    end
    Sample $\boldsymbol{\beta} | \alpha, \gamma, s_1, \ldots, s_n$;
    for $k \leftarrow 1$ **to** K **do**
        for $m \leftarrow 1$ **to** N_{MH} **do**
           Sample $\sqsubset'_k \sim q(\sqsubset'_k | \sqsubset_k)$;
           Sample $t \sim \text{Uniform}(0,1)$;
           $r \leftarrow \min\left(1, \frac{p(\sqsubset' | \boldsymbol{X}^k) q(\sqsubset | \sqsubset')}{p(\sqsubset | \boldsymbol{X}^k) q(\sqsubset' | \sqsubset)}\right)$;
           if $t < r$ **then**
               $\sqsubset_k \leftarrow \sqsubset'_k$
           end
        end
    end
end

Algorithm 1: MCMC sampler for HDP-HMM Bayesian Networks

3 Applications

3.1 Example – simulated data

To evaluate the efficacy of our method we generated simulated data from three different Bayesian network structures and interleaved the data points into a single time series. Applying our methodology to this data we then attempted to recover the hidden state sequence.

Three different Bayesian Networks of 10 nodes each having random structures and parameters were used, with the restriction that each node had at most two parents. Such a restriction is realistic for real world biological networks and reduces the computational complexity of the Bayesian network inference as the number of potential sets of parents of each node is greatly reduced by constraining the search. A total of 100 data points were used, consisting of a sequence of 25 generated by the first network, 25 by a second network structure, another 25 from a third network structure and finally a further 25 data points generated by the second network structure.

The Gibbs sampling MCMC scheme outlined in algorithm 1 was applied over 500 iterations after a 1000 iteration burn in, with 100 MCMC iterations of the Bayesian network order sampler run on each network structure between each Gibbs sweep.

In figure 2 we show a comparison of the true hidden state sequence with the state sequences for the 500 samples from the Gibbs sampler. Our method perfectly recreates the original hidden state sequence, correctly identifying that the network structure is the same between two separate segments of the sequence of data points.
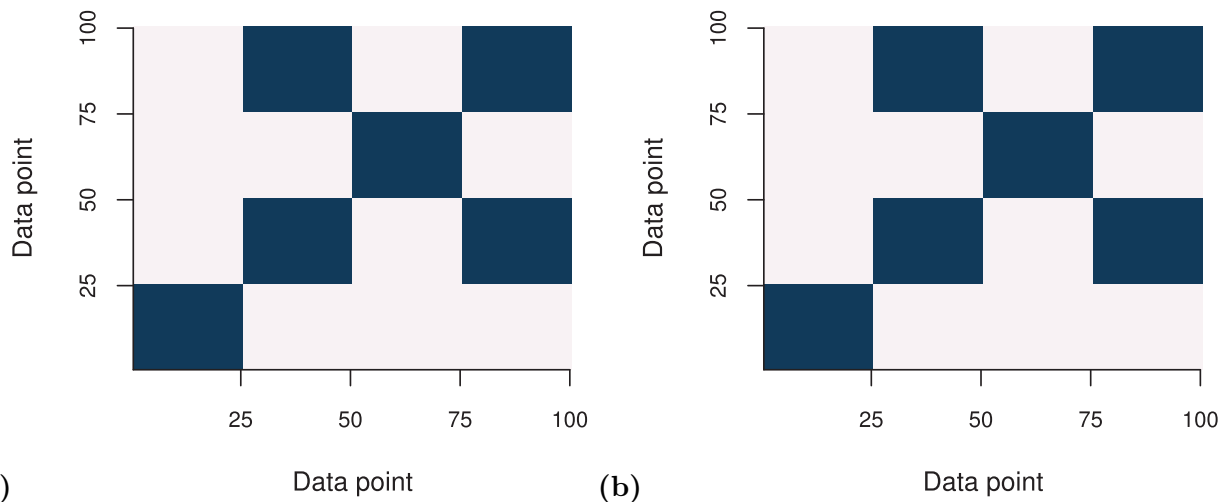
Figure 2: **(a)** Counts of the number of the 500 posterior samples in which pairs of data points where assigned to the same state. **(b)** The same illustration for the true hidden state sequence.

3.2 *D. melanogaster* midgut development gene expression data

Applying our method to real world gene expression data, we took the publicly available gene expression data set of Li and White [15], as stored in the Gene Expression Omnibus database [2]. This data set gives tissue specific expression levels for genes in *D. melanogaster* midgut at time points before and after pupariam formation, taken at 11 time points. A subset of genes to analyse was chosen by selecting genes having the highest variance across the time series, using the genefilter R package in Bioconductor `www.bioconductor.org` [8, 19]. This resulted in a data set of 23 genes at 11 time points. This allows us to apply our approach without having to consider the additional issues arising the 'large-p-small-n' problem.

The results shown in figure 3 identify two regions of the time series having different network structures, with a change occuring after the 0 hour time point at which pupariam formation occurs. This suggests that a different structure of regulatory interactions is at work during the midgut development after the pupariam formation begins. The networks inferred for each of the different states are shown in figure 4, illustrating a clear change between differing network structures.

3.3 Transcriptome of starch metabolism during *A. thaliana* diurnal cycle

We have also analysed the gene expression data set of Smith *et al.* [24], as included in the GeneNet [21] R package [19]. The data set consists of expression levels for 800 genes encoding enzymes involved in starch synthesis and in conversion of starch to maltose and Glc, at 11 time points over 12 hours, transitioning from dark to light. The first 5 time points were collected during a dark period after which a switch to a light period was made, with time points spaced so that expression is measured at 0, 1, 2, 4 and 8 hours after the switch to the dark period, and the same intervals after the switch to the light period [24], as well as a final 24 hour time point at the switch back to the dark period. A reduced subset of the 800 genes in the data set was selected using the genefilter R package, as described previously, giving a subset of 40 genes that were analysed using our method.

In figure 5 we show the results generated by our method, clearly indicating two distinct phases within the time series. It appears that one phase is detected from 1 to 12 hours, with a second phase inferred between 13 and 24 hours, that is also represented at the initial time point. This is consistent with the design of the experiment, as a change in expression would perhaps not be expected to be
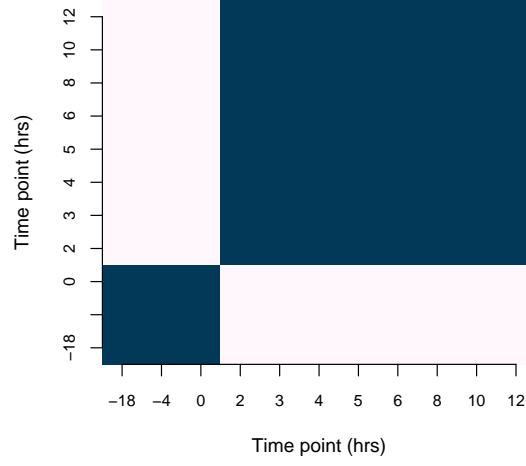
7

Figure 3: Posterior distribution of states at each time point inferred by our method applied to the *D. melanogaster* midgut development expression data [15]. States are represented by colours, and frequencies of their appearance for each time point in the posterior samples is plotted.
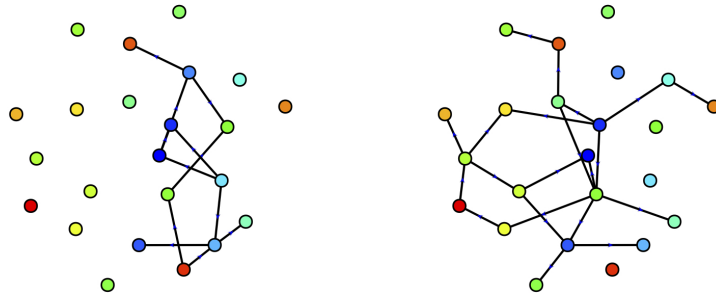

Figure 4: Sampled Bayesian Network structures for the two states inferred by our method applied to the *D. melanogaster* cell cycle expression data [15].
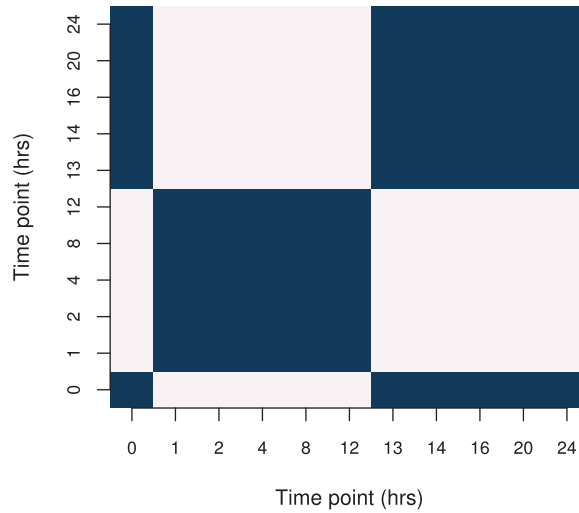
8

Figure 5: Posterior distribution of states at each time point inferred by our method applied to the *A. thaliana* diurnal cycle expression data [24]. States are represented by colours, and frequencies of their appearance for each time point in the posterior samples is plotted.
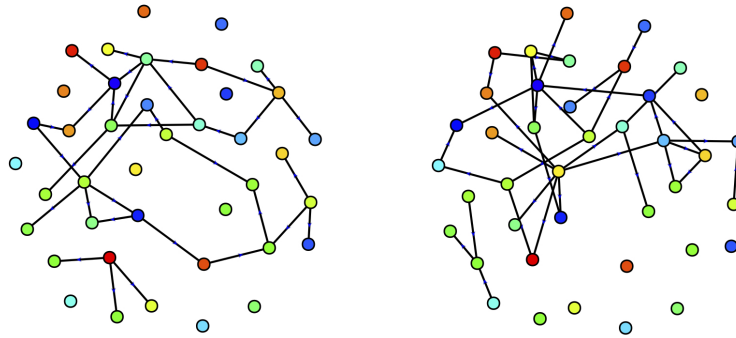

Figure 6: Sampled Bayesian Network structures for the two states inferred by our method applied to the *A. thaliana* diurnal cycle expression data [24].

observed immediately at the point at which the switch between light and dark takes place, but rather later at a subsequent time point, as is observed here. Since the 24 hour time point was taken under the same conditions as the initial time point, one would expect these two time points to be grouped together using our method. The networks inferred for the two different phases, shown in figure 6, again demonstrate a clear change in the network structure, with the two networks having distinct topologies.

4 Discussion

From our simulated data it is clear that the HDP-HMM Bayesian Network sampler we have constructed accurately infers the hidden state sequences governing Bayesian Networks that capture how the regulatory organisation of a biological system changes with time. The accuracy of our method on test data suggests that it will perform well on real world data sets, whilst the existence of more

9

sophisticated and demonstrably more efficient samplers indicates that there is room for even further improvement and computational efficiency. For example the beam sampler of Van Gael *et al.* [27] and the Hierarchical Chinese Restaurant Process (HCRP) formalism of Makino *et al.* [17] have better mixing and perform better than standard Gibbs samplers, especially on time series such as those we examine here where neighbouring states are likely to be correlated.

We remark that it is essential to consider the fluid nature of regulatory network structures when inferring networks from data sets where such change is likely. Performing an analysis on data using a model with a fixed network structure, when it is known that the network structure will change due to some stimulus, is inherently incorrect, and thus will introduce unnecessary bias into the results. Whilst it may be possible to infer correct results from an incorrect model, it would not seem wise to rely on such approaches when alternatives exist.

Our methodology crucially accounts for the sequential nature of the data, something that has previously been overlooked [10, 11], but we feel is crucial to the modelling of gene expression time series data sets. Furthermore our methodology has an advantage over changepoint models that data may be shared between distinct segments of the time series sharing the same hidden state when inferring the network structure – something that is explicitly represented in our model, but generally omitted in changepoint models. This is especially import in gene expression data analysis where time points are a scarce and valuable resource.

The versatility of the HDP-HMM means that our methodology is applicable not only to time series data where the underlying process is divided into distinct contiguous segments, as would be expected in gene regulatory networks, but also to processes describable by a Markov process, for example rapidly changing between a sequence of hidden states with some underlying transition mechanism. Thus it may be of use for other problems of network inference in systems biology outside of the area of sequential gene expression time series data, or in other fields where networks that change with time are encountered.

Finally whilst other methods may require manual specification of an appropriate prior distribution on the number of possible states, taking a nonparametric approach allows our prior distribution to naturally expand to explain the observed data as the size and complexity of the data grows. Bayesian nonparametric methods demonstrably outperform regular priors in a variety of applications and we have shown here their potential in modelling hidden variables in theoretical systems biology.

References

[1] Beal, M., Ghahramani, Z. and Rasmussen, C.E. The infinite hidden Markov model. *Advances in neural information Processing Systems*, 2002.

[2] Edgar, R., Domrachev, M. and Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002.

[3] Ellis, B. and Wong, W.H. Learning causal Bayesian network structures from experimental data. *Journal of the American Statistical Association*, 103(482):778–789, 2008.

[4] Fox, E.B., Sudderth, E.B., Jordan, M.I. and Willsky, A.S. An HDP-HMM for systems with state persistence. In *ICML '08: Proceedings of the 25th international conference on Machine learning*. ACM, 2008.

[5] Fox, E.B., Sudderth, E.B., Jordan, M.I. and Willsky, A.S. A sticky HDP-HMM with application to speaker diarization. *arXiv.org*, stat.ME, 2009.

[6] Friedman, N. and Koller, D. SpringerLink - Machine Learning, Volume 50, Numbers 1-2. *Machine Learning*, 50(1/2):95–125, 2003.

[7] Geiger, D. and Heckerman, D. Parameter Priors for Directed Acyclic Graphical Models and the Characterization of Several Probability Distributions. *Annals of statistics*, 2002.

[8] Gentleman, R., Carey, V., Huber, W. and Hahne, F. *genefilter: genefilter: methods for filtering genes from microarray experiments*, 2007. R package version 1.34.0.

[9] Grzegorczyk, M. and Husmeier, D. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. *Machine Learning*, 71(2-3), 2008.

[10] Grzegorczyk, M., Husmeier, D., Edwards, K.D., Ghazal, P. and Millar, A.J. Modelling non-stationary gene regulatory processes with a non-homogeneous Bayesian network and the allocation sampler. *Bioinformatics (Oxford, England)*, 24(18):2071–2078, 2008.

[11] Ickstadt, K. Nonparametric Bayesian Networks (with discussion). In J. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, editors, *Bayesian Statistics 9*, pages 135–155. Oxford University Press, 2011.

[12] Koski, T. and Noble, J. *Bayesian Networks: An Introduction (Wiley Series in Probability and Statistics)*. Wiley, 1st edition, 2009.

[13] Lèbre, S. Inferring dynamic genetic networks with low order independencies. *Statistical applications in genetics and molecular biology*, 8(1):Article 9–, 2009.

[14] Lèbre, S., Becq, J., Devaux, F., Stumpf, M.P.H. and Lelandais, G. Statistical inference of the time-varying structure of gene-regulation networks. *BMC systems biology*, 4:130–, 2010.

[15] Li, T.R. and White, K.P. Tissue-specific gene expression and ecdysone-regulated genomic networks in Drosophila. *Developmental cell*, 5(1):59–72, 2003.

[16] Madigan, D., York, J. and Allard, D. Bayesian Graphical Models for Discrete Data. *International Statistical Review*, 63(2):215–232, 1995.

[17] Makino, T., Takei, S., Sato, I. and Mochihashi, D. Restricted Collapsed Draw: Accurate Sampling for Hierarchical Chinese Restaurant Process Hidden Markov Models. *arXiv.org*, stat.ML, 2011.

[18] Opgen-Rhein, R. and Strimmer, K. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology*, 1:37, 2007.

[19] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[20] Rodriguez, A., Lenkoski, A. and Dobra, A. Sparse covariance estimation in heterogeneous samples. *arXiv.org*, stat.ME:4208, 2010.

[21] Schafer, J. and Opgen-Rhein, R. Reverse engineering genetic networks using the GeneNet package, 2006.

[22] Schäfer, J. and Strimmer, K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics (Oxford, England)*, 21(6):754–764, 2005.

[23] Sethuraman, J. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

[24] Smith, S., Fulton, D., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S. *et al.* Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis leaves. *Plant Physiology*, 136(1):2687–2699, 2004.

[25] Teh, Y.W. and Jordan, M.I. Hierarchical Bayesian nonparametric models with applications. In *Bayesian nonparametrics*, pages 158–207. Cambridge Univ. Press, Cambridge, 2010.

[26] Teh, Y., Jordan, M. and Beal, M. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 2006.

[27] Van Gael, J., Saatci, Y., Teh, Y.W. and Ghahramani, Z. Beam sampling for the infinite hidden Markov model. In *ICML '08: Proceedings of the 25th international conference on Machine learning.* ACM, 2008.